Paper:

An Unsupervised Learning Method for Industrial Anomaly Detection Based on VAE-LSTM Hybrid Model

Hengshan Zong[1], Haolong Zhang[1], Ning Jia[1], Maohua Gong[1], Hengmin Dong[1], Zeyu Jiao[2,*]

[1]China Aerospace Academy of Systems Science and Engineering, Beijing, China

[2]Guangdong Key Laboratory of Modern Control Technology, Institute of Intelligent Manufacturing, Guangdong Academy of Sciences, Guang Zhou, China

E-mail:zy.jiao@giim.ac.cn

[Received 00/00/00; accepted 00/00/00]

Anomaly detection is one of the most important tasks in industrial production, and it is a crucial aspect for both safety and efficiency of modern process industries. However, due to the high-dimensional characteristics of time series perception data and the difficulty of data labeling in actual production, there is still a lack of effective anomaly detection methods in industrial scenarios. To solve the above challenges, this research proposes an unsupervised learning method based on a hybrid model of Variational Autoencoder (VAE) and Long Short-term Memory (LSTM). First of all, high-dimensional industrial data is processed by VAE, not only achieves dimensionality reduction and feature extraction, but also reduces the impact of noise. Then the LSTM network is exploited to mine the temporal features of the industrial data and predict the subsequent change trend. Finally, when the difference between the predicted data and the actual measured data exceeds a certain threshold, the production process can be considered abnormal.

Keywords: Industrial anomaly detection, Unsupervised learning, Hybrid model, VAE, LSTM

1. Introduction

In large-scale industrial production, the equipment is always operated under certain conditions to ensure that the products produced have the same quality and function. However, the quality of products may be affected by equipment aging and other abnormal conditions, even cause serious accidents. Therefore, it is very important to monitor the industrial production process and detect anomaly conditions to ensure product quality and carry out preventive maintenance to reduce losses [5]. With the rapid development of industrial Internet, modern industrial manufacturing system has realized the perception and recording of the production status, environment and process, and accumulated a large amount of industrial data through sensors, controllers, intelligent instruments and other monitoring equipment. [8].

It has become a mainstream method to monitor the running state of equipment and detect potential abnormal events with the help of time-series sensing data collected by industrial sensors [14]. On the one hand, industrial production is a continuous behavior. On the other hand, the monitoring equipment of industrial production is highly diversified, and the accumulated industrial data is a typical multi-dimensional time-series data. Therefore, anomaly detection based on multi-dimensional time-series data has received great attention in the field of industrial production. Liu et al. [15] proposed a new deep anomaly detection framework based on on-device federated learning for the time series data perception of the Industrial Internet of Things, and adopted an attention-mechanism-based convolutional neural networked long and short time memory (AMCNN-LSTM) model to accurately detect anomalies. Hsieh et al. [9] studied a case of anomaly detection in intelligent manufacturing. Using the real data collected from the sensing equipment of the factory production line, in order to overcome the limited and irregular anomaly patterns in the multivariable sensor data, an unsupervised real-time anomaly detection algorithm based on the LSTM autoencoder was proposed.Nearly 90% accuracy was achieved in both accuracy and recall rates. Quatrini et al. [18] proposed a two-step industrial process anomaly detection method using machine learning classification algorithm. Starting with the real-time collection of process data, the first step identifies the ongoing process phase, the second step classifies the input data as "expected", "warning" or "critical" and validates the proposed anomaly detection approach in a real-world case study in the pharmaceutical industry.

Despite lots efforts have been paid, anomaly detection of multidimensional time-series data for industrial production is a very challenging task, mainly due to: 1) There is potential correlation and mutual influence between different dimensions of data, which makes it more difficult to detect and identify abnormal patterns [27]; 2) Industrial big data has a series of characteristics, such as large volume, multisource heterogeneous, strong dynamic, etc., making data processing more difficult [20]; 3) The analysis and processing of industrial sensing data rely on a large amount of expert experience, and large-scale annotation of data is a costly work, which also makes the method of supervised learning difficult to achieve [24].

Aiming to solve the above challenges, this study proposed a VAE and LSTM based unsupervised learning method for industrial anomaly detection. First, VAE is leveraged to encode high-dimensional data to reduce the influence of sensor noise on monitoring results while achieving data reduction and feature selection. The LSTM is then exploited to predict subsequent trends based on historical data, and abnormal events are considered when the predicted value and the actual measured value exceed a certain threshold. In the constructed VAE-LSTM hybrid model, VAE first encodes the original input information, and then decodes it. The difference between the decoded information and the original input information is used to guide the network training. This process does not rely on manual annotation of data. In addition, LSTM only predicts the trend of subsequent data based on the historical data itself, and also does not require manual annotation. Thus, the method proposed in this study is an unsupervised method, which can effectively avoid the consumption of human resources and cost caused by the need for expert annotation. The contributions of this study can be summarized as:

- An unsupervised learning method based on VAE-LSTM hybrid model for industrial anomaly detection is proposed.
- VAE is utilized to map high dimensional data into low vector space and to dig for potential correlation.
- LSTM is exploited to mine temporal features of data based on historical data and predict the trend of subsequent data.
- The experimental results show that the proposed method outperforms the existing methods.

2. Related works

2.1. Anomaly detection

The existing research shows that it is feasible to detect abnormal events based on high-dimensional industrial time series data analysis. Kanawaday and Sane [12] deployed a variety of sensors in production equipment to collect timing data, and use this data to predict failures and optimize the manufacturing process. They reviewed the development history of anomaly detection and divided anomaly detection methods into traditional detection methods, supervised learning based detection methods and unsupervised learning based detection methods.

Traditional anomaly detection methods, such as KNN [1], local anomaly factor LOF algorithm [3], and connection-based anomaly point factor COF algorithm [25], distinguish normal and abnormal data by the similarity between samples. However, these methods have the limitations of high computational complexity and high rate of missed detection, and are not suitable for high-dimensional data.

To deal with these problems, researchers have introduced supervised machine learning. Griffin et al. [7] proposed an approach based on neural networks and decision tree, which is used to detect anomalies in multiple processing processes, opening the door for control implementation. Nanduri et al. [17] built an application of a recurrent neural network (RNN) with LTSM and a gated recurrent unit (GRU) structure to detect abnormal events from multivariate time series data collected from aircraft flight data recorder (FDR) or flight operation quality assurance (FOQA) data. Unknown anomalous patterns in the data can be detected through semi-supervised or unsupervised learning. Janssens [10] proposed a feature learning model for rotating machinery state monitoring based on convolutional neural network. The objective of this method is to learn the useful features of bearing fault detection from the data itself. The results show that the performance of the feature learning system based on Convolutional Neural Network (CNN) is obviously better than the classical feature engineering method based on manual feature design and random forest classifier, and the accuracy rate reaches 87.25%. With supervised learning method, the model can learn the difference between normal data and abnormal data, which can effectively improve the accuracy of anomaly detection. Any monitoring method will inevitably require some or all of the data annotated information.However, in practical application, a common difficulty is that the historical data is completely or mostly not annotated and the annotation cost is too high.

To solve these problems, researchers have proposed some unsupervised anomaly detection methods [2, 4, 6, 11, 13, 16, 22]. Amruthnath et al. [2] applied several unsupervised learning methods to anomaly detection, such as K-means and fuzzy C-means clustering.Anomaly detection by these methods can be defined as the process of identifying deviations from standard behavior, which is also the most common detection method of unsupervised anomaly detection. Diez et al. [4] proposed an anomaly detection method based on One-class SVM, which obtained the anomaly score based on the distance between the sample and the separation hyperplane, so as to detect the anomaly in the sensor data. Joshi et al. [11] proposed anomaly detection based on Hidden Markov Model (HMM), which established HMM by extracting features and calculating the anomaly probability in the state sequence generated by the model. Traditional unsupervised anomaly detection methods all need to solve a practical and basic problem, that is, to determine its structure, that is, to find a way to adjust model parameters, but there is no good way to solve this problem. Feng et al. [6] proposed a dynamic autoregressive comprehensive moving average model (DARIMA), which established a prediction model through the correlation between time series data and judged the abnormal results according to the threshold value. It was mainly applied to short-term prediction and could extract the short-term time series dependence relationship in the data. Serdio et al. [22] proposed an anomaly detection method for coal plants in power plants based on continuous learning fuzzy model and dynamic residual analysis, and no anomaly labeling was required with this model. Kingma [13] applied the variational autoencoder (VAE) to analyze the residuals of the reconstructed data and the source data to detect abnormalities through the reconstruction data. Lu et al. [16] constructed a stacked autoencoder and its variant VAE based on a series of autoencoders. The use of a layered method for deep network structure learning is conducive to obtaining high-level feature representations of complex sensory signals. And these high-level feature representations can be processed as the input of subsequent fault classifiers through an unsupervised learning model. However, the traditional unsupervised learning method has a poor effect on the high-dimensional time-series data, while in the real industrial production process, most of the data obtained are characterized by high dimension and high dynamic.

Traditional stacked autoencoders can process high-dimensional data, but the effect is not good in time series data, while LSTM network can effectively extract the time series characteristics of data.By combining the characteristics of VAE and LSTM network, the network architecture based on the LSTM and VAE unsupervised detection model autoencoder can process multi-dimensional nonlinear data and learn the normal behavior of unlabeled data sets.

2.2. Evaluation metric

As for anomaly detection, the recognition rate and misjudgment rate of abnormal events are the focus of attention, the following indexes can be used to evaluate the advantages and disadvantages of the model. The recognition rate of exceptional events can be expressed as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

The recognition rate for normal conditions is calculated by the following formula

$$\text{Recall} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (2)$$

The overall performance of the model can be measured by F1-score which can be obtained as follows

$$\text{F1} - \text{score} = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

where TP is an abnormal event that is accurately identified, FP is a normal condition that is judged as an abnormal event, TN is a normal condition that is accurately identified and FN is an abnormal event that is misjudged as a normal condition.

3. Method

3.1. Overall scheme

The overall scheme of the proposed VAE-LSTM hybird method is shown in Fig. 1. First , the collected time series sensing data is divided into a series of small segments of equal length by the sliding window. The sensing data is then entered into VAE, which is encoded by the encoder and reconstructed by the decoder. Finally, LSTM is used to predict the trend of subsequent data changes based on VAE encoded vectors.

3.2. VAE-LSTM hybird model

For the high-dimensional sensing information in the production process, the VAE model is constructed to extract the multi-dimensional information from the adaptation. Suppose that the input data of the model is the high-dimensional sensing information x collected in the production process, and the encoding vector z is generated after the encoder $q_\phi(z \mid x)$. The whole encoding and decoding process can be expressed as:

$$z \sim \text{Enc}(x) = q_\phi(z \mid x) \quad (4)$$

$$x \sim \text{Dec}(z) = p_\theta(x \mid z) \quad (5)$$

where the encoder learns the distribution of input data, maps the input data to the mean μ and standard deviation σ of the data distribution, and samples it in the standard normal distribution to generate the potential coding vector z.

$$z = \mu + \sigma \odot \varepsilon \quad (6)$$

where $\odot$ is the Hadamard product. The decoder reconstructs the underlying variable z to generate the reconstructed sample. In VAE, the model is optimized by maximizing the evidence lower bound function

$$L_b = Eq_\phi(z \mid x) \log p_\theta(x \mid z) - D_{KL}\left(q_\phi(z \mid x) \mid p_\theta(z)\right) \quad (7)$$

where $Eq_\phi(z \mid x)$ is the logarithmic likelihood estimation of a posteriori probability of x, represents the reconstruction quality. $D_{KL}(*)$ is KL divergence, which is used to measure the difference between the approximate posterior distribution and the unit Gaussian distribution.

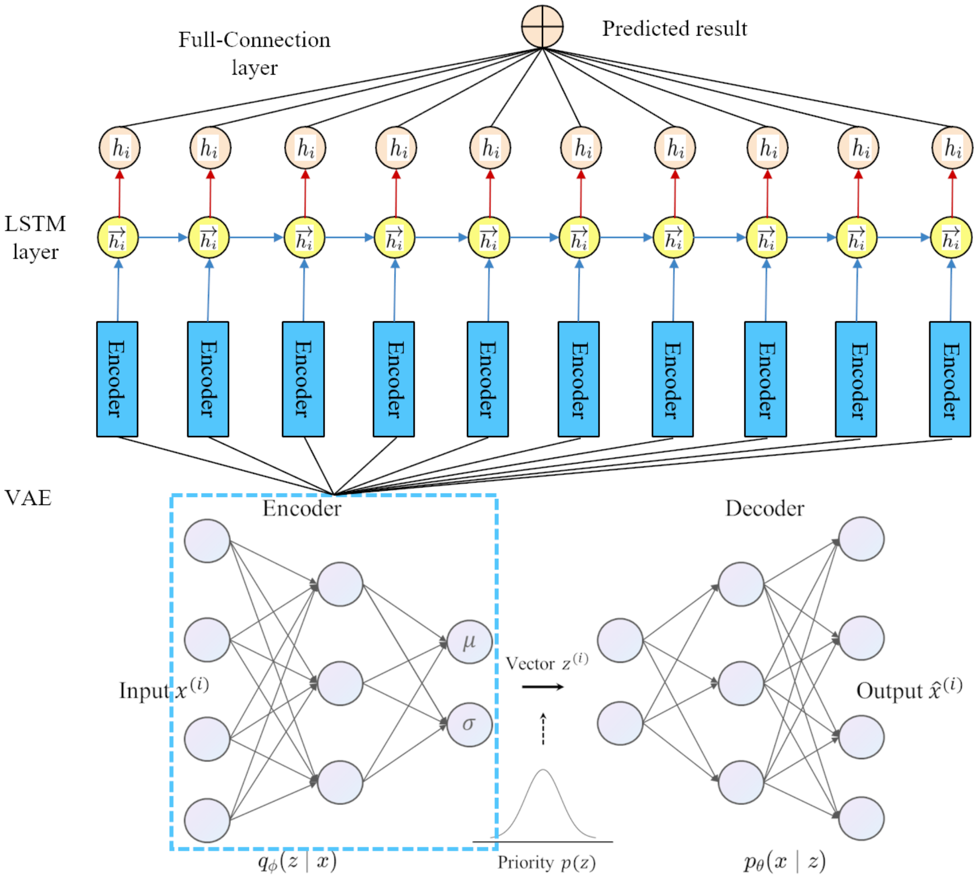

Fig. 1. Overall scheme of the proposed method

In the whole process of VAE coding and decoding, the high-dimensional sensing data are gradually reduced and selected through multi-layer neural network, which can effectively screen out the factors that have a greater impact on the representation of the state of the production process and ignore the secondary effects.Therefore, the model is robust to noises, missing values and errors that are widely existed in actual production. After that, LSTM unit is used to model the coding vector on the time sequence and mine the time sequence correlation relationship, so as to realize the extraction and encoding of the time sequence information and generate the prediction result embedded vector in the potential semantic space of corresponding time sequence features. The architecture of the LSTM is illustrated in Fig. 2. The basic structure of LSTM is made up of cells, and a cell is composed of several gates, including forget gate, memory gate and output gate.

The task of the forget gate is to take a long term memory C_{t-1} (the output from the previous cell module) and decide which parts of the formula to keep and forget, which can be expressed as

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (8)$$

The purpose of a memegate is to determine what new information is stored in a cellular state. The

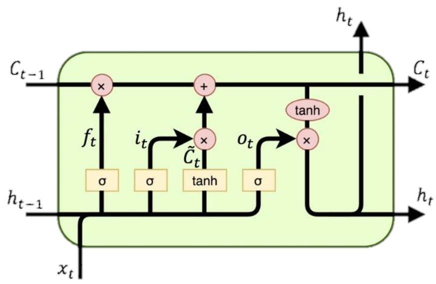

Fig. 2. Schematic diagram of the LSTM network structure

memory gate consists of two parts: 1) the Sigmoid layer, which determines what values need to be updated; 2) the Tanh layer, create a new candidate value vector, generate candidate memory. The process can be calculated by

$$i_t = \sigma\left(W_t \cdot [h_{t-1}, x_t] + b_i\right) \quad (9)$$

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \quad (10)$$

At this point, the old cell state C_{t-1} has been updated to obtain the value of C_t :

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (11)$$

Finally, depending on the state of the cell, the output of the LSTM can be defined. A Sigmiod function is used to determine which part of the cell state needs to be output, and then the cell state is processed through the tanh layer, and the two are multiplied to get the final desired output information.

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \quad (12)$$

$$h_t = o_t * \tanh\left(C_t\right) \quad (13)$$

4. Experimental results

4.1. Dataset

In this paper, an experiment is conducted based on a real data set from a human-computer integration workshop. The workshop is a continuous industrial production process, during which all kinds of anomalies (such as mechanical failure, personnel misoperation, etc.) may occur. The data set contains data collected from multiple sensors on the production line at a frequency of 0.5Hz. Therefore, the sensor data collected is a multi-dimensional time series data, including 18,398 samples. In addition, the dataset provides annotated information of exception events, including 124 moments when exceptions were annotated. The fields of the data are as follows:

- t: The timestamp represents the time when the data was collected
- $x_1 - x_{27}$: Characteristic variables represent readings taken from different sensors, including device status and operating status, etc.

It should be noted that, for the consideration of information security and other perspectives, all the data are desensitized. It is not the data of the original sensor, but contains the basic characteristics of the running state of the device.

4.2. Experimental details

In the experiment, 10 moments of industrial sensor parameters are used as input to predict the parameter value at the 11th moment. The initial learning rate was 1.5×10^{-3}, and the decay rate was set as 0.1 per 1000 iterations. SGD was selected as the training optimizer, and the momentum was 0.9. The activation function of the hidden layer was the rectified linear unit (ReLu) function and the activation function of the output layer was the linear function. All the experiments were conducted on the Pytorch deep learning framework and open source libraries such as mmdetection, python-opencv, and skimage under the ubuntu 16.04 operating system and Visual Studio Code environment. The hardware configurations were CPU: Intel(R) Core i7-9750H; memory: 16.0 GB; GPU: Nvidia Geforce GTX 1660 Ti. The programming language was python 3.6 and the integrated development environment was Anaconda 3.

Meanwhile, the method proposed in this study was compared with some existing studies, including Autoregressive Integrated Moving Average model (ARIMA) [23], One-Class Support Vector Machine (OC-SVM), Interpreting Random Forests (iForest) [26], AutoEncoder and LSTM (AE-LSTM) [19] and Deep Autoregressive Recurrent (DeepAR) [21]. Fig. 3 visualizes the reconstruction results of VAE in a certain dimension and Fig. 4 shows the performance of the proposed model for the prediction of data from a certain sensor.

Table 1. Experimental results of the comparison experiment

Model	Precision(%)	Recall(%)	F1-score
ARIMA	67.31	45.78	54.50
OC-SVM	35.49	62.04	45.15
iForest	71.82	66.6	69.11
AE-LSTM	94.55	92.49	93.51
DeepAR	98.07	96.73	97.40
VAE-LSTM	97.56	99.31	98.43

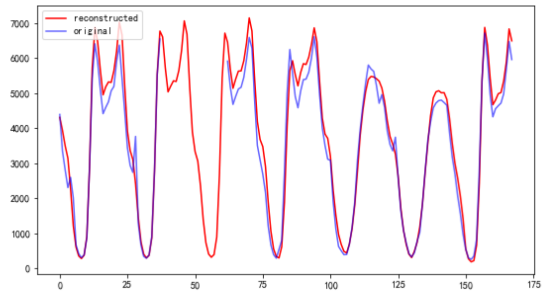

Fig. 3. Difference between reconstructed result and the original input

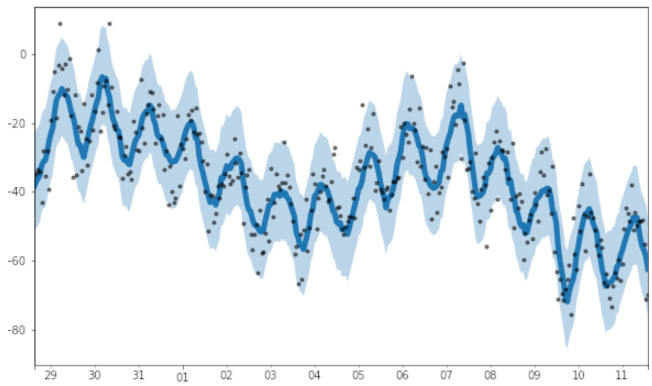

Fig. 4. Predicted result of the hybrid model on a single sensor

5. Conclusion

In this study, an unsupervised industrial anomaly detection method based on VAE and LSTM was proposed.First, the input data is preprocessed into periods of the same length. VAE is then utilized to encode the input high-dimensional data to form a vector in the latent semantic space. Next, LSTM is used to mine the temporal features in the latent semantic space and predict the subsequent trends of sensor values based on this. Finally, when the difference between the predicted results and the actual measured results exceeds a certain threshold, the system is considered abnormal. The experimental results show that the method proposed in this study outperform the existing research, and can realize the accurate detection of anomalies, which is of great significance.

Acknowledgements

This work is supported by the financial support from GDAS' Project of Science and Technology Development (Grant No. 2021GDASYL-20210103090), Guangzhou Science and Technology Plan Project (Grant No. 202007040007).

References:

[1] H. A. Abu Alfeilat, A. B. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat, H. S. Eyal Salman, and V. S. Prasath, "Effects of distance measure choice on k-nearest neighbor classifier performance: a review", Big data, 7(4):221–248, 2019.

[2] N. Amruthnath and T. Gupta, "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance", In 2018 5th International Conference on Industrial Engineering and Applications (ICIEA), pp. 355–361. IEEE, 2018.

[3] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers", In Proceedings of the 2000 ACM SIGMOD international conference on Management of data, pp. 93–104, 2000.

[4] A. Diez-Olivan, J. A. Pagan, N. L. D. Khoa, R. Sanz, and B. Sierra, "Kernel-based support vector machines for automated health status assessment in monitoring sensor data", The International Journal of Advanced Manufacturing Technology, 95(1):327–340, 2018.

[5] C. Feng, T. Li, and D. Chana, "Multi-level anomaly detection in industrial control systems via package signatures and LSTM networks", In 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 261–272. IEEE, 2017.

[6] T. Feng, Z. Du, Y. Sun, J. Wei, J. Bi, and J. Liu, "Real-time anomaly detection of short-time-scale gwac survey light curves", In 2017 IEEE International Congress on Big Data (BigData Congress), pp. 224–231. IEEE, 2017.

[7] J. M. Griffin, A. J. Doberti, V. Hernández, N. A. Miranda, and M. A. Vélez, "Multiple classification of the force and acceleration signals extracted during multiple machine processes: part 1 intelligent classification from an anomaly perspective", The International Journal of Advanced Manufacturing Technology, 93(1):811–823, 2017.

[8] D. H. Hoang and H. D. Nguyen, "A PCA-based method for IoT network traffic anomaly detection", In 2018 20th International conference on advanced communication technology (ICACT), pp. 381–386. IEEE, 2018.

[9] R.-J. Hsieh, J. Chou, and C.-H. Ho, "Unsupervised online anomaly detection on multivariate sensing time series data for smart manufacturing", In 2019 IEEE 12th Conference on Service-Oriented Computing and Applications (SOCA), pp. 90–97. IEEE, 2019.

[10] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. V. d. Walle, and S. Van Hoecke, "Convolutional neural network based fault detection for rotating machinery", Journal of Sound and Vibration, 377:331–345, 2016.

[11] S. S. Joshi and V. V. Phoha, "Investigating hidden Markov models capabilities in anomaly detection", In Proceedings of the 43rd annual Southeast regional conference-Volume 1, pp. 98–103, 2005.

[12] A. Kanawaday and A. Sane, "Machine learning for predictive maintenance of industrial machines using IoT sensor data", In 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 87–90. IEEE, 2017.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational bayes", arXiv preprint arXiv:1312.6114, 2013.

[14] S. Lin, R. Clark, R. Birke, S. Schönborn, N. Trigoni, and S. Roberts, "Anomaly Detection for Time Series Using VAE-LSTM Hybrid Model", In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4322–4326. IEEE, 2020.

[15] Y. Liu, S. Garg, J. Nie, Y. Zhang, Z. Xiong, J. Kang, and M. S. Hossain, "Deep anomaly detection for time-series data in industrial iot: A communication-efficient on-device federated learning approach", IEEE Internet of Things Journal, 2020.

[16] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification", Signal Processing, 130:377–388, 2017.

[17] A. Nanduri and L. Sherry, "Anomaly detection in aircraft data using Recurrent Neural Networks (RNN)", In 2016 Integrated Communications Navigation and Surveillance (ICNS), pp. 5C2–1. Ieee, 2016.

[18] E. Quatrini, F. Costantino, G. Di Gravio, and R. Patriarca, "Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities", Journal of Manufacturing Systems, 56:117–132, 2020.

[19] Z. Que, Y. Liu, C. Guo, X. Niu, Y. Zhu, and W. Luk, "Real-time Anomaly Detection for Flight Testing using AutoEncoder and LSTM", In 2019 International Conference on Field-Programmable Technology (ICFPT), pp. 379–382. IEEE, 2019.

[20] T. Rieger, S. Regier, I. Stengel, and N. L. Clarke, "Fast Predictive Maintenance in Industrial Internet of Things (IIoT) with Deep Learning (DL): A Review.", In CERC, pp. 69–80, 2019.

[21] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks", International Journal of Forecasting, 36(3):1181–1191, 2020.

[22] F. Serdio, E. Lughofer, K. Pichler, T. Buchegger, and H. Efendic, "Residual-based fault detection using soft computing techniques for condition monitoring at rolling mills", Information Sciences, 259:304–320, 2014.

[23] H. Song, S. Sui, Q. Han, H. Zhang, and Z. Yang, "Autoregressive integrated moving average model–based secure data aggregation for wireless sensor networks", International Journal of Distributed Sensor Networks, 16(3):1550147720912958, 2020.

[24] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning", In International conference on artificial neural networks, pp. 270–279. Springer, 2018.

[25] J. Tang, Z. Chen, A. W.-C. Fu, and D. W. Cheung, "Enhancing effectiveness of outlier detections for low density patterns", In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 535–548. Springer, 2002.

[26] X. Zhao, Y. Wu, D. L. Lee, and W. Cui, "iForest: Interpreting random forests via visual analytics", IEEE transactions on visualization and computer graphics, 25(1):407–416, 2018.

[27] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data", Statistical Analysis and Data Mining: The ASA Data Science Journal, 5(5):363–387, 2012.